**To be published in the proceedings of
4th International Symposium on
Spatial Accuracy Assessment in
Natural Resources and Environmental Sciences
July 2000**

A Case Study in GIS-Based Environmental Model Validation Using Earthquake-Induced Landslide Hazard

Scott B. Miles
U.S. Geological Survey, Department of Geography, Box 353550, University of Washington, Seattle, WA 98103, USA, smiles@swave.wr.usgs.gov

David K. Keefer
U.S. Geological Survey, 345 Middlefield Rd., MS 977, Menlo Park, CA USA, dkeefer@usgs.gov

Timothy L. Nyerges
Department of Geography, Box 353550, University of Washington, Seattle, WA 98103, USA, nyerges@u.washington.edu

Keywords: validation, environmental modelling

Abstract

We present a case-study of validating a particular model of regional earthquake-induced landslide hazard using the common approach of comparing model output to empirical observations. We hope to illuminate several issues and obstacles with regards to model validation that have not been directly discussed in the validation literature. Issues of model integration with GIS, which have not been specifically related to model validation in the environmental modelling literature, are also raised. Based on the details of the case study, we argue that one cannot arrive at an absolute conclusion regarding the validity of a model by simply comparing model output to empirical observations. Model output may compare well simply because of a serendipitous combination of data and decisions. We argue that the adjective "valid" only has contextual meaning. It only applies to the network of relationships that emerge out of a particular decision-making context. Such a network includes decision-makers, stakeholders, and modellers, as well as, the study area, data, model performance, software, etc. Without this network, the process of validation and evaluation is meaningless. This process must be a social conversation among decision-makers, stakeholders, and modellers in order to assess model usefulness and establish trust in the network.

1. Introduction

The efforts of the many researchers who have sought to demonstrate the significant utility of environmental modelling with GIS are beginning to pay off. GIS-based environmental models and the various forms of their results are now widely used within an array of decision-making contexts. Decisions based GIS-based environmental models can have widespread social, political, and economic impact. Thus, as the popularity and use of GIS-based environmental models in decision-making settings continues to grow, assessing the legitimacy or soundness of such models becomes critical to their development, selection, and use. This process of assessment is usually associated with the term "validation." There are no specific approaches to model validation that are commonly accepted across science and engineering. However, this process is universally associated with checking the output of a model against a set of empirical observations, standard, or statistical criteria, upon which the model in question is given a stamp of "valid" or "not valid."

Several pan-discipline studies have critically examined model validation from the perspective of epistemology (e.g., Barlas and Carpenter, 1990, Oreskes et al., 1994) or meaning and procedure (e.g., Rykiel, 1996). These excellent studies are conceptual and general, but are not placed within the context of a specific application. We think there is significant value in examining the process of validating a particular GIS-based environmental model, with which we are familiar. This allows us to 1) provide detail that may otherwise be blurred through generalisation, 2) comment on a broad range of topics and, 3) encourage a degree of reflexivity on the part of modellers.

Thus, we present a case-study of validating a model for regional earthquake-induced landslide hazard. Using the detail of the case study, we hope to illuminate several issues and obstacles with regard to model validation that have not been directly discussed in the validation literature. Similarly, issues and obstacles with respect to model integration with GIS, which have not been specifically related to model validation in the environmental modelling literature, will be raised. While we have made efforts to relate the specifics of our case study to environmental models in general, it is unavoidable that many issues are specific to the particular model. However, it is in the particulars that many of the broader issues become more obvious. The disciplines of GIS have long discussed issues of data quality. With this paper, we hope to advocate and broaden the discussion surrounding model validation. This discussion is critical in dealing with the growing success of our efforts to popularise GIS-based environmental models for decision and policy-making. We begin our discussion by introducing the subject of our case-study. We then present the validation of our model in five relatively distinct steps. We conclude by extracting basic themes from the case study and making a recommendation regarding the issue of model validation.

2. Earthquake-Induced Landslide Hazard

In many earthquakes, triggered landslides have accounted for most of the economic losses or casualties. Perhaps the most devastating example is the death of more than 120,000 people during the an **M**=7.8 earthquake in China in 1920 (Wang and Xu, 1984). Landslides caused by the 1989 Loma Prieta, CA earthquake (**M**=6.9) damaged at least 200 residences, caused at least $30 million in damage and blocked highways in the epicentral region for several weeks (Keefer, 1998). Most recently, the 1999 Chi-Chi, Taiwan earthquake (**M**=7.6) caused over 7000 landslides. Considering the widespread and costly effects of seismically triggered landslides, hazard zonation can play a critical role in pre-event planning and post-event mitigation.

Several models have been developed for analysing earthquake-induced landslides (see Ho and

Miles, 1997 and Miles and Keefer, 1999 for a review). Several popular models are based on Newmark's displacement method (1965), which was originally developed to analyse man-made embankments. It was later extended to analyse natural slopes (Wilson and Keefer, 1983). The method models a potential landslide as rigid friction-block, having a known critical acceleration, resting on an inclined plane. The major assumptions of this analogy are 1) the slope is rigid and perfectly plastic, 2) a well-defined slip surface exists, 3) that shear strength remains constant during shaking. These assumptions are not representative of general landslide behaviour (Kramer, 1996).

Newmark's method calculates the cumulative displacement of the friction-block as it is subjected to an earthquake acceleration time-history. This is done by double-integrating those parts of the earthquake time-history that exceed the critical acceleration (Wilson and Keefer, 1985). Newmark (1965) defined the following relationship to calculate critical acceleration:

$$a_c = (FS - 1)\sin\alpha \qquad (1)$$

where FS is the static factor of safety of the slope and α is the thrust angle of the landslide block. For GIS-based slope-stability analysis, the nearly universal approach to calculating the static factor of safety is the infinite slope model, which requires the treatment of each pixel (or other morphological unit) as an infinitely-long slope.

$$FS = \frac{c'}{\gamma d \sin\alpha} + \frac{\tan\phi'}{\tan\alpha} + \frac{m\gamma_w \tan\phi'}{\gamma \tan\alpha} \qquad (2)$$

where c' is the effective cohesion, ϕ' is the effective angle of internal friction, γ is the material unit weight, γ_w is the unit weight of water, α is the angle of the slope from the horizontal, d is the normal depth to the failure surface, and m is a ratio of d indicating the relative location of the ground water table.

Various simplified models derived from Newmark's method have enjoyed popularity for application with GIS. Most of these simplified models are based on numerical regression of results obtained by applying conventional Newmark's method with a generic set of inputs. The most popular of these simplified models was developed by Jibson (1993), which has subsequently been modified (Jibson et al., 1998).

3. A Case Study in Validation

We present the process of validating the simplified Newmark model of Jibson et al. (1998). We have chosen to validate this model because of its popularity, ease of implementation using GIS, and our familiarity with it (see Miles and Ho, 1999a, Miles and Keefer, 1999). The model calculates Newmark displacements (D_n) as a function of critical acceleration (a_c) and Arias intensity (I_a), which is a descriptor of earthquake shaking intensity. The equation takes the following form.

$$\log D_N = 1.521 \log I_a - 1.993 \log a_c - 1.546 \qquad (3)$$

In the following sections, we break the validation process into five steps: 1) obtaining empirical observations, 2) obtaining model inputs, 3) applying the model, 4) comparing model outputs to observations, and 5) assessing model validity and performing sensitivity analysis.

4. Empirical Observations

Before we can compare model output with reality, we must have an inventory of landslides caused by some earthquake. Unfortunately, there are only a handful of such inventories around the world associated with large earthquakes. One particular inventory was gathered after the Loma Prieta earthquake (**M**=6.9) (Keefer, 1998). For this study, we focused on the landslides within the Laurel quadrangle, which contains the epicenter of the earthquake. The inventory includes attributes of landslide type (i.e., failure mode), geologic material, and estimated volume of displaced material. Except for a few very large landslides, each landslide is represented by a single co-ordinate location. Because of dense vegetation cover, landslides could not be mapped using aerial-photographs. Because the inventory was gathered from traverses in vehicles and on foot, inventory completeness is somewhat uncertain (Jibson, personal communication).

5. Input data

Having empirical observations, we need to collect, estimate, or generate input data for the model. For the simplified Newmark model, three types of data are needed: hill slope-angles, the Arias intensity for each slope, and strength properties of slope materials. Hill slope-angles can be calculated from a digital elevation model (DEM), which is easy to obtain in many parts of the world. Regardless of availability, issues related to DEMs, such as resolution, error, and form are important considerations. However, a common attitude of modellers is to leave consideration of these issues to the makers of DEMs. But, these issues can have notable effects on the results of GIS-based models. Hartshorne (1997) demonstrated that GIS-based slope-stability models are highly sensitive to the different slope geometries produced by different DEM resolutions and forms. For this study, we used a 30-meter USGS format DEM generated from a 1:24,000 topographic map of the quadrangle.

The model requires regional values for Arias intensity, which is typically calculated directly from earthquake acceleration time-histories. Such a requirement can pose a significant obstacle due to scarcity of data (Miles and Ho, 1999a). One solution is to calculate Arias intensity for locations where time-histories exist and then interpolate. Another solution is to use an empirically derived attenuation relationship. As is common in most areas with earthquake-induced landslides, only a few earthquakes records of the Loma Prieta earthquake exist for our study area. In addressing this issue, some studies assume a single value or earthquake record (e.g., McCrink and Real, 1996). However, we opted to use an attenuation relationship (see also, Miles and Keefer, 1999). We used the relationship of Wilson (1993), which takes the following form for estimating mean Arias intensity (I_a).

$$\log I_a = M - 2\log\sqrt{R^2 + h^2} - 3.99 \tag{4}$$

The parameter R is the minimum horizontal distance to the vertical projection of the fault plane and h is a correction factor that defaults to 7.5 km. In modelling the ground-motions from the Loma Prieta earthquake, we used the fault-rupture model of Marshall et al. (1991) from which to calculate *R*. By using the attenuation relationship and the fault-rupture model, we have brought other models into the fray that require independent validation. Undoubtedly, if a different equation or fault-rupture model is used, results of our Newmark analysis will be different. One may argue that it would be better to find an earthquake (and related landslide inventory) that had a larger, more complete set of earthquake records so as to remove uncertainty associated with the attenuation relationship from the validation process. However, the greatest demand for modelling the regional effects of earthquakes is often in regions where large earthquakes have not recently occurred, for example the Oakland, CA area (see Miles and Keefer, 1999). In such cases, earthquake strong-motion data will be lacking and a similar solution as the one employed for this case study will have to be used. So the use of an attenuation relationship may be more appropriate and relevant.

The last type of data required are the engineering strength properties of the regional geotechical materials. Newmark's method was originally developed for analysing man-made embankments, which are constructed out of materials with relatively well-known properties. For natural slopes, this is rarely ever the case. Determining strength properties is often fraught with difficulty, uncertainty, and high-costs. Properties are either determined from intensive sampling and laboratory testing or from in-situ tests. Of course, sampling and in-situ testing quickly becomes impractical at larger geographic scales.

Coping strategies for regional analysis include using "book values" and expert knowledge or compilation of shear strength test data from disparate sources such as county engineering reports or sparse field testing. The latter approach has many strong proponents because of its apparent objectivity and defensability. However, this approach also requires a high degree of modeller judgement and abstraction. For example, test data from several dissimilar sources and test types may be obtained and grouped based on geologic formation. Various data records may be excluded based on modeller-defined criteria before further manipulation. The remaining test data for each formations are likely averaged in some way to yield a single "representative" value. Interestingly, Keefer (in press) found that landslide concentration (number of landslides per square kilometre) from the Loma Prieta earthquake did not correlate well with strength properties that had been specifically compiled for regional application of Newmark's method. Landslide concentration did however exhibit relatively strong correlation with linguistic descriptions of the geologic formations.

Most approaches for regional shear strength characterisation involve the assignment of a single value to large spatial units: typically geologic formations or soil survey units. This approach first makes the tenuous assumption that there is no spatial variability within spatial units. Second, this method makes the assumption that there is a correlation between the particular definition of spatial unit and shear strength. In the case of geologic formations, the latter assumption is poor because the formations are not classified based on shear strength, but rather the age and lithology of the predominant rocks. A formation often consists of many rock types of varying conditions. The makeup of geologic units also varies with map scale; large-scale maps are not universally available. Similarly, soil survey maps were not created as a support for shear strength estimates.

For this study, we started with the shear strength database compiled for a previous study of earthquake-induced landslides in the Laurel quadrangle (McCrink and Real, 1996). We chose to only use records associated with peak strength because the database includes few records on residual strength. As noted, we do not believe that compiled data is necessarily any better than expert knowledge or "book values." However for validation, we wanted to use what most people would consider as the best available data. Mean values were assigned to formations of a geologic map of the quadrangle (Wentworth, 1993), although the dataset primarily reflects soil properties. If a formation had no data associated with it, values for similar formations were averaged and assigned to the unit. We assumed that each landslide failed at constant depth of 3.33 meters. Although an obvious simplification, this value is considered representative (Keefer, 1984). Lastly, because the Loma Prieta earthquake occurred at the end of a dry summer season during a four year drought (Keefer, 1998), dry conditions were assumed and the groundwater depth parameter of Eq.(2) was set to zero.

6. Model implementation

Having obtained empirical observations and, arguably, the best available input data, the next step is to implement the model within some computing environment. The topic of integrating environmental models with GIS is well-documented, but there are no standard guidelines or well-used recipes (Miles and Ho, 1999b). The task of integration is much more of an art, being specific to the characteristics of a particular model and intended application of the model.

Different implementation strategies will likely result in different model output (Miles and Ho, 1999b). For example, the use of GIS-specific data structures will lead to particular distortions of which we may not be aware or, rather, choose to ignore (Burrough and Frank, 1995). Environmental models are often implemented using a raster data model; thus, we must choose a resolution for implementation. When elevation or slope is a parameter of the model, many modellers do not view this as a decision and simply default to the resolution of the DEM being used. However, this passive decision, which many view as objective, induces a high degree of uncertainty if other data are not of equal or higher resolution. For seismic landslide modelling and other similar cases, geologic or soil survey units that are assigned attribute values are orders of magnitude coarser than typical DEM resolutions. Transforming geologic units to a higher resolution requires deaggregation. Deaggregation is troublesome because unit averages have no information regarding point values within the unit. There is no unique solution to deaggregation; thus uncertainty arises regarding values derived through deaggregation (Heuvelink and Pebesma, 1999). In light of the uncertainties associated with deaggregation, the use of higher and higher resolution DEMs may not improve model performance unless the modelled system is entirely slope dependent.

Considering the issue of deaggregation, we should have run the analysis at the resolution of the coarsest data. Thus, we would calculate an average slope and average Arias intensity for each geologic unit, then calculate a single Newmark displacement for each unit. Regardless, we made the common choice of adopting the resolution of our DEM. For better or worse, modellers and end-users seem more apt to prefer the uncertainties of deaggregation to the unintuitive concept of a single slope or ground-shaking intensity value for a very large area that may not even be contiguous.

7. Comparison of output and observations

Having implemented the model and obtained output, the next step is to compare model output to the empirical observations. Using the example of our case study, we can see that this is not always a simple matter. The results of a Newmark analysis take the form of a numerical index having units of displacement (usually centimetres). Determining the displacement for all landslides caused by a particular earthquake is often impossible, even with the assistance of remote sensing. The landslide inventory for the Loma Prieta earthquake does not contain information regarding displacement, except for a few large landslides. Even if such information was available, a single displacement value may not be representative of a disrupted or rotational landslide. Because of this and other issues, for regional analysis, Newmark displacement is typically considered as just an index of slope performance (Jibson et al., 1998).

Viewing Newmark displacement as a performance index, rather than real-world deformation, does not remove the disjunction between the model output and the landslide inventory. One solution to this disjunction is to convert Newmark displacement into a probability of landslide using a standard probability model, calibrated using actual data (Jibson et al., 1998). To date, the only other solution is the use of some critical Newmark displacement, above which complete failure of the slope (i.e., landslide) is assumed. This concept has no ties to the original Newmark method (1965), and there is little consensus regarding what is a representative critical displacement. Typical values used, which are all largely anecdotal, include 2, 5 or 10 cm, but can be as high as 300 cm (see Matasovic, 1991). Respective values are commonly related to specific landslide failure modes, which can make *a priori* selection of a value difficult. No study has directly investigated the validity of the concept for regional modelling. However, the treatment of Newmark displacement at regional scales as a performance index, rather than a measure of real-world deformation, seems inconsistent with the use of a critical displacement, which is based on real-world deformation.

Regardless of its problems, critical Newmark displacement is the only means of transforming the continuous displacement values into a binary assessment of landslide/no-landslide. Thus, we must choose a critical Newmark displacement value to compare model output to the landslide inventory. Obviously, the value we choose will have a large effect on the outcome of the comparison. Although not representative of the failure mode of all Loma Prieta landslides, we assumed a value of 10 cm because it is one of the more widely cited values.

After transforming model output, we still must transform the landslide inventory from point locations and polygons (for the few large landslides) to pixels in a grid. Thus we must choose a resolution to represent the landslides. The obvious choice is to use the same resolution used for model implementation. But again, this transformation of resolution imposes a degree of uncertainty (Heuvelink and Pebesma, 1999). The resolution chosen may not reflect the varying sizes of respective landslides. For example, if we rasterise the point locations of of the landslide inventory at 10 meters, there are 412 landslide pixels, but at 30 meters there are only 404. Obviously, some small landslides fall within a single 30-meter pixel. Considering the majority of landslides in the Laurel quadrangle have a volume less than 100 m^3, we feel comfortable using a 30-meter resolution — conveniently, the resolution of model implementation.

Thus, having the inventory rasterised and the model output in terms of landslide/no-landslide, we can compare the two to assess model performance. We can determine how many actual landslides were "captured" by the Newmark model by summing the total number of landslide pixels that correspond with pixels having Newmark displacement greater than 10 cm — our assumed critical Newmark displacement. In this way, we find that only 69 landslides or 17% of the total 404 landslides were captured. Of course, there are pixels with Newmark displacements greater than 10 cm that do not correspond to landslides in the inventory. In fact, there are almost 18,000 pixels or 10.5% of the total quadrangle area (154 km^2) with Newmark displacement greater than 10 cm.

Two other studies have made similar comparisons between the results of a Newmark model and the Loma Prieta landslides. Mankelow and Murphy (1998) were able to capture 49% of the landslides, using the model of Jibson (1993), in their particular study area, which was a 32 km^2 area within the Laurel quadrangle. However, they used a critical Newmark displacement of 5 cm. They reported that 18% of the total study area was predicted to have Newmark displacements greater than 5 cm. After trying 22 combinations of critical Newmark displacements, strength values, and groundwater conditions, McCrink and Real (1996) presented the case they felt lead to the best performance of their implementation of conventional Newmark's method for the Laurel quadrangle. The best case, with a critical Newmark displacement of 5 cm, managed to capture 84% of the landslides, while 50% of the quadrangle was predicted to have Newmark displacement greater than 5 cm.

The question of optimum balance between maximising the number of captured landslides and minimising the overall area that is wrongly modelled to be effected by landslides brings up the issue of which is more important to predict: landslide location or relative landslide concentration. It of course would be easy to capture all of the Loma Prieta landslides by simply painting the entire Laurel quadrangle red. Thus, we suggest that landslide concentration is of first order importance. For example, a relative ranking of geologic units based on modelled landslide concentration would be highly useful and, in fact, is needed before predictions regarding individual landslide location can have meaning.

Therefore, we wanted to assess the results from the simplified Newmark model with respect to the landslide concentrations within each geologic unit in the Laurel quadrangle. Table 1 (column 3) lists, in descending order, the number of landslide pixels per square kilometre calculated for each unit from the Loma Prieta inventory. Calculated landslide concentration will vary with the support area used in determining concentrations. For example, Keefer (in press) calculated that the Purisima formation had the highest concentration (2.03 landslides/km^2) for a 15 quadrangle area encompassing the Loma Prieta epicenter. Whereas, for this study of the Laurel quadrangle, the Purisima formation has only the ninth highest concentration (0.44 landslides/km^2).

Using the simplified Newmark model and the same critical Newmark displacement, we calculated the modelled landslide concentration within each geologic unit. Table 1 (column 4) lists the modelled concentrations in comparison to the actual concentrations. Listed also are the original strength properties in modelling the concentrations (column 4). The modelled concentrations for a few of the geologic units are significantly greater than concentrations calculated from the landslide inventory, while several units were modelled to have zero landslides even though, in reality, landslides occurred. However, of the seven geologic units that have the highest actual landslide concentration, five of the units were also modelled to be among the seven most hazardous.

Geologic Unit	Percent total area	Actual Landslide pixels/km^2	Modelled Landslide pixels/km^2	Original Strength Estimates φ (degrees)	Original Strength Estimates c' (kPa)	Back-calculated Strength φ (degrees)	Back-calculated Strength c' (kPa)
Diabase, db	0.004%	85.71	0.00	32.2	35.1	30.0	35.0
Landslide Deposit, Qls	5.85%	10.92	75.78	14.7	10.3	24.4	52.5
Butano Mudstone, Tbm	1.29%	7.92	0.00	19.0	62.2	27.6	50.0
Rices Mudstone, Tsr	4.08%	6.79	8.01	32.5	36.1	32.3	47.5
Vaqueros Sandstone, Tv	7.52%	2.36	18.94	30.5	31.9	35.0	45.0
Butano Formation, Tb	1.82%	1.85	100.00	31.4	14.9	35.1	47.3
Marine Shale/Sandstone, Tme	0.19%	1.31	1.97	23.0	19.2	34.7	50.0
Lompico Sandstone, Tlo	1.75%	0.49	0.00	36.0	65.2	29.2	30.0
Purisima Formation, Tp	55.93%	0.44	0.05	33.7	30.9	29.5	42.0
Alluvium, Qal	3.57%	0.29	0.17	35.0	23.9	33.6	34.5
Lambert Shale, Tla	2.65%	0.28	72.91	26.4	20.5	38.3	39.8
Santa Margarita Sandstone, Tsm	0.45%	0.27	0.00	29.0	31.6	29.7	39.0
Twobar Shale Member, Tst	0.70%	0.26	100.00	27.7	17.0	35.4	50.0
San Lorenzo Formation, Tsl	0.24%	0.26	2.30	30.6	40.9	35.0	50.5
Salinian Basement Granite, Kgrm	0.28%	0.22	0.00	32.2	35.1	32.5	37.3

Table 1, Results of sensitivity analysis

8. Assessment of validity

Having compared model output with empirical observations, one can ask whether the comparison confirms or rejects the validity of the particular model. In our case, we should now be able to make some assessment as to the validity of the simplified Newmark model. Looking at the results of our comparison, the model certainly did not perform perfectly. It did not capture all of the landslides, nor did it correctly predict the absolute or relative landslide concentrations for each geologic unit. But, some landslides were captured and some of the geologic units that had the highest concentration of landslides were predicted to have high landslide concentrations relative to other geologic units.

Obviously, the results of the comparison is directly related to the obvious uncertainties in the input data and landslide inventory, in addition to uncertainties imposed through transformations, such as rasterisation, and arbitrary choices, such as assuming a critical Newmark displacement of 10 cm. It is common to try to deal with the uncertainty associated with input data. One strategy used to investigate the effect of input data on model output is a sensitivity analysis. This involves varying inputs to see how model output is effected. If a large effect is observed, it may be justifiable to attempt to improve the data quality or simply assume that parameter estimates are highly uncertain.

We chose to investigate the sensitivity of the simplified Newmark model to variations in material strength properties. Rather than randomly vary or perturb the strength property inputs, we decided to determine whether or not, by modifying the original input data, we could match the predicted landslide concentrations for each geologic unit to the actual concentrations. With the assistance of a interactive macro, we were successfully able to match predicted and actual landslide concentrations by arbitrarily adjusting the values for friction angle and effective cohesion (see Eq.(2)). The back-calculated values are listed for each geologic unit in Table 1 (last column). Noting the difference between landslide concentrations calculated using the original data and the concentrations calculated with the back-calculated values (i.e., the actual concentrations), we see that the model is very sensitive to the strength property inputs. Thus, in practice, we would either decide to attempt to improve estimates or, more practically, advocate consideration of high uncertainty in any related decision context.

The results of this sensitivity analysis are interesting for another reason. We acknowledged and identified many areas of uncertainty in the previous four sections. Further, the model investigated was a simplification of the original Newmark model, which in turn certainly does not account for all governing factors for earthquake-induced landslides. Even so, we were able to produce model predictions that exactly matched empirical observations through a serendipitous combination of inputs and assumptions. This seems to raise doubt regarding whether the validity of the model can be absolutely confirmed or rejected through comparison of model output and empirical observations.

Thus, the question still remains whether the model is valid or not. McCrink and Real (1996) answer by stating that "the Newmark method, as refined by [their] investigation, is a viable approach to mapping earthquake-induced landslides on a regional basis." This is a relatively strong, positive statement. However, the word “viable” (interpretted as possible or workable) seems carefully chosen and not intended as a synonym for valid. Mankelow and Murphy (1998) make a similar statement regarding the results of their study, but qualify it as relative to another particular earthquake-induced landslide model. As described above, the results of the comparisons performed as part of these two studies, were similar to the comparison results obtained in this study. Model output from both studies managed to capture more landslides than in our case, but conversely results of both model implementations wrongly predicted a greater percentage of area as being effected by landslides. Neither study used landslide concentration as a basis for comparison. Even so, we have three validation studies of basically the same model in basically the same area, that produce very different measures of performance. It would appear that even with three validation studies of Newmark's displacement method for regional analysis, the question of validity is still left unanswered. Considering, we argue that we cannot arrive at an absolute conclusion regarding the validity of the simplified Newmark model by simply comparing model output to the Loma Prieta landslide inventory.

10.Towards a Contextual View of Validity

By presenting this case study, we hope to have illuminated many issues regarding various stages involved in implementation and validation of the particular model. Many of these issues directly translate to the domain of other environmental models, while some issues relate indirectly. However, there are two broad themes that we would like to extract from the particulars of the case study. First, in illuminating several issues and obstacles along the way of implementing (and validating) the model using GIS, it is clear that many decisions are involved, many of which are commonly made passively, but made nonetheless. We feel that it is important to realise this and, thus, make decisions explicit, being aware of the potential effect and possible alternatives. For example, if it is decided that the model will be implemented at the resolution of the particular DEM being used, actively make this decision so as to be aware that by doing so additional uncertainty will be induced if deaggregation of coarser units is required. The argument of whether a particular decision is objective or subjective is moot if we make every effort to be aware of the possible implications of each decision involved in model implementation.

The second theme that we would like to put forward is that "valid" and "not valid" are not a attributes of environmental models, GIS-based or otherwise. Foremost, validity does not refer to the truth or correctness of a model; a model is neither true or not-true. A model of the environment is a statement about an open system; thus, the truth of this statement cannot be established, even by comparing model output to empirical observations (see Oreskes et al., 1994). We feel that this is demonstrated well by the case study. Even accepting that validation is not about establishing truth, a model is still liable to be stamped as "valid" if it meets or exceeds some modeller-specified performance criteria (Rykiel, 1996). But as we tried to demonstrate with the case study, model output may compare well with criteria simply because of a serendipitous combination of data and decisions. By making other choices regarding data, transformations, GIS-model integration, etc., model output may no longer meet the criteria. This is similar to attempting to validate a model for a different area.

Thus, we agree with Oreskes (1998) who, in espousing evaluation over validation, stresses that we should never describe any model as valid. Just as the attribute "distance" applies to the relationship between, for example, Seattle and Amsterdam, "valid" and "not valid" only have contextual meaning. The adjectives only apply to the network of relationships that emerge out of a particular decision-making context. Such a network includes humans, such as decision-makers, stakeholders, and modellers, as well as, non-humans, such as the study area, data, model performance, software, etc. Without this network, the process of validation and evaluation is meaningless. This process must be a social conversation (Barlas and Carpenter, 1990) among decision-makers, stakeholders, and modellers in order to assess usefulness and usability. The result of this conversation is to establish trust in the network and to place a stamp of "valid" or "useful" or "valuable" on the associated relationships, not the model. Because of the complexity and dynamism of any such network, the label "valid" is always a matter of degree and will always be short-lived. Thus, validation and evaluation must be a never-ending process. This will prevent burying the uncertainties and nuances that arise within particular modelling contexts.

In conclusion, publication of comparisons of model output with empirical observations should not be viewed as a requirement of modellers. While such comparisons have significant utility in demonstrating the potential use of a model and illuminating issues of practical application, they are only indirectly useful in determining validity. Rather, we need to continue our discussion on validation and evaluation to develop means of facilitating effective conversations and interactions among elements of decision-making networks.

Acknowledgements
The authors would like to thank Nick Chrisman (University of Washington) and Ray Wilson (U.S. Geological Survey) for their prompt reviews and thoughtful comments.

References

– *Barlas, Y. and Carpenter, S., 1990*, Philosophical roots of model validation: two paradigms, System Dynamics Review, 6(2), pp. 148-146.

- *Burrough, P.A., Frank, A.U., 1995*, Concepts and paradigms in spatial information: are current geographical information systems truly generic?, International Journal of GIS, 9(2), pp. 101-116.
- *Heuvelink, G.B.M., Pebesma, E.J., 1999*, Spatial aggregation and soil process modelling, Geoderma, 89, pp. 47-65.
- *Hartshorne, J., 1997*, Assessing the influence of digital terrain model characteristics on tropics slope stability analysis, in: Innovations in GIS 4, Kemp, Z., Eds., Taylor & Francis, pp. 198-214.
- *Ho, C.L., Miles, S.B., 1997*, Deterministic Zonation of Seismic Slope Instability: An Application of GIS, in: ASCE Geotechnical Special Publication No. 67: Spatial Analysis in Soil Dynamics and Earthquake Engineering, pp. 87-102.
- *Jibson, R.W., Harp, E.L., Michael, J.A., 1998*, A Method for Producing Digital Probabilistic Seismic Landslide Hazard Maps: An Example from the Los Angeles, California area, U.S. Geological Survey Open-File Report 98-113, 17 pp., 2 pl.
- *Jibson, R.W., 1993*, Predicting Earthquake-Induced Landslide Displacements Using Newmark's Sliding Block Analysis, Transportation Research Record 1411, pp. 9-17.
- *Keefer, D.K. [in press]*, Statistical analysis of an earthquake-induced landslide distribution — The 1989 Loma Prieta, California event, Engineering Geology.
- *Keefer, David K., ed., 1998*, The Loma Prieta, California, Earthquake of October 17, 1989 ¾ landslides, U.S. Geological Survey Professional Paper 1551-C, 183 pp.
- *Keefer, David K., 1984*, Landslides Caused by Earthquakes, Geological Society of America Bulletin, 95, pp. 406-421.
- *Kramer, Steven L., 1996*, Geotechnical Earthquake Engineering, Prentice Hall, New Jersey, 653 p.
- *Mankelow, J.M., Murphy, W., 1998*, Using GIS in the probabilistic assessment of earthquake triggered landslide hazards, Journal of Earthquake Engineering, 2(4), pp. 593-623.
- *Marshall, G.A., Stein, R.S., Thatcher, W., 1991*, Faulting geometry and slip from coseismic elevation changes; the 18 October, 1989 Loma Prieta, California, earthquake, Seismological Society of America Bulletin, 81, pp. 1660-1693.
- *Matasovic, N., 1991*, Selection of method for seismic slope stability analysis, Proc.: 2nd International Conference on Recent Advances in Geotechnical Earthquake Engineering and Soil Dynamics, pp. 1057-1062
- *McCrink, T.P., Real, C.R., 1996*, Evaluation of the Newmark Method for Mapping Earthquake-Induced Landslide Hazards in the Laurel 7.5’ Quadrangle, Santa Cruz County, California: Final Technical Report for U.S. Geological Survey Contract 143-93-G-2334, 47 p.
- *Miles, S.B., Ho, C.L., 1999a*, Rigorous Landslide Hazard Zonation using newmark's Method and Stochastic Ground Motion Simulation, Soil Dynamics and Earthquake Engineering, 18(4), pp. 305-323.
- *Miles, S.B., Ho, C.L., 1999b*, Applications and issues of GIS in civil engineering modeling, Journal of Computing in Civil Engineering, 13(3), pp. 144-152.
- *Miles, S.B., Keefer, D.K., 1999*, Comparison of Seismic Slope-Performance Models Case Study of the Oakland East Quadrangle, CA, U.S. Geological Survey Open-File Report, 38 p.
- *Newmark, N.M., 1965*, Effects of Earthquakes on Dams and Embankments, Geotechnique, 15, pp. 139-160.

- *Oreskes, N., 1998*, Evaluation (Not Validation) of quantitative models, Environmental Health Perspectives, 106, Sup. 6, pp. 1453-1459.
- *Oreskes, N., Shrader-Frechete, K., Belitz, K., 1994*, Verification, validation, and confirmation of numerical models in the earth sciences, Science, 263(5147), pp. 641-646.
- *Rykiel, E.J., 1996*, Testing ecological models: the meaning of validation, Ecological Modelling, 90, pp. 229-244.
- *Wang, W., Xu, B., 1984*, Brief introduction of landslides in loess in China, 4th International Symposium on Landslides, Toronto, Proceedings, 1, pp. 197-207.
- *Wentworth, C.M., 1993*, General distribution of geologic materials in the southern San Francisco Bay region, California: a digital map database, U.S. Geological Survey Open-File Report 93-693, 12 p.
- *Wilson, R.C., Keefer, D.K., 1985*, Predicting Areal Limits of Earthquake-Induced Landsliding, U.S. Geological Survey Professional Paper 1360, pp. 317-494.
- *Wilson, R.C., Keefer, D.K., 1983*, Dynamic Analysis of a Slope Failure From the 6 August, 1979 Coyote Lake, California, earthquake, Seismological Society of America Bulletin, 73(3), pp. 863-877.
- *Wilson, R.C., 1993*, Relation of Arias Intensity to Magnitude and Distance in California, U.S. Geological Survey Open-File Report 93-556, 42 p.